Exploring Prime Areas in Bandung for Tourism Business Opportunity

By: Ryan Varyan

April 1, 2020

1. Introduction

1.1. Background

Bandung is the capital city of West Java - Indonesia. In the early 20th century the Dutch inhabitants of Bandung demanded the establishment of a municipality (gamete), which was granted in 1906, and Bandung gradually developed into a resort city for plantation owners. Luxurious hotels, restaurants, cafés, and European boutiques were opened, leading the city to be nicknamed Parijs van Java (Dutch: "The Paris of Java") (Ref. Wikipedia).

Bandung tourism industry increases exponentially as the Airport accommodates more international flights. This attracts investors locally and internationally to establish their business in Bandung. However, due to language barrier and lack of information of tourist destination, foreigner who wish to invest may find it difficult to understand the regulation and best location before they decide to open a business related to tourism.

This report is created for potential investors to determine the best location for their business based on demographic data of neighborhood in Bandung.

1.2. Problem

Bandung tourism industry increases exponentially as the Airport accommodates more international flights. This attracts investors locally and internationally to establish their business in Bandung. However, due to language barrier and lack of information of tourist destination, foreigner who wish to invest may find it difficult to understand the regulation and best location before they decide to open a business.

1.3. Interest

This report is created for potential investors to determine the best location for their business based on demographic data of neighborhood in Bandung.

2. Data

To perform the analysis, the following data were obtained:

2.1. List of the districts of Bandung

List of districts were obtained from Wikipedia.

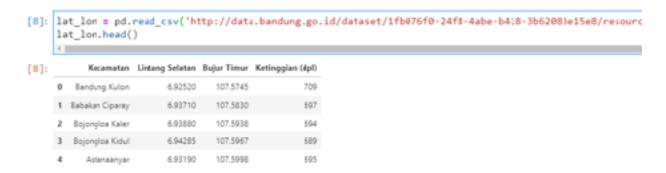https://id.wikipedia.org/wiki/Daftar_kecamatan_dan_kelurahan_di_Kota_Bandung

Web scrapping tool, BeautifulSoup, was utilized to extract the districts in Bandung from Wikipedia. The code and table of districts are shown below.

```
[2]: source = requests.get('https://id.wikipedia.org/wiki/Daftar_kecamatan_dan_kelurahan_di_Kota_Bandung').text
     soup=BeautifulSoup(source, 'html5lib')

[3]: tables = soup.find_all('table')

[4]: df = pd.read_html(str(tables[0]), flavor='bs4')[0]

     df
```

[4]:

	Kode Kemendagri	Kecamatan	Jumlah Kelurahan	Daftar Kelurahan
0	32.73.06	Andir	6	Campaka Ciroyom Dunguscariang Garuda Kebonjeru...
1	32.73.10	Astana Anyar	6	Cibadak Karanganyar Karasak Nyengseret Panjuna...
2	32.73.20	Antapani	4	Antapani Kidul Antapani Kulon Antapani Tengah ...
3	32.73.24	Arcamanik	4	Cisaranten Bina Harapan Cisaranten Endah Cisar...
4	32.73.03	Babakan Ciparay	6	Babakan Babakanciparay Cirangrang Margahayu Ut...
5	32.73.21	Bandung Kidul	4	Batununggal Kujangsari Mengger Wates
6	32.73.15	Bandung Kulon	8	Caringin Cibuntu Cigondewah Kaler Cigondewah K...
7	32.73.09	Bandung Wetan	3	Cihapit Citarum Tamansari
8	32.73.12	Batununggal	8	Binong Cibangkong Gumuruh Kacapiring Kebongeda...
9	32.73.04	Bojongloa Kaler	5	Babakan Asih Babakan Tarogong Jamika Kopo Suka...
10	32.73.17	Bojongloa Kidul	6	Cibaduyut Cibaduyut Kidul Cibaduyut Wetan Kebo...
11	32.73.22	Buahbatu	4	Cijawura Jatisari Margasari Sekejati
12	32.73.18	Cibeunying Kaler	4	Cigadung Cihaurgeulis Neglasari Sukaluyu
13	32.73.14	Cibeunying Kidul	6	Cicadas Cikutra Padasuka Pasirlayung Sukamaju ...
14	32.73.25	Cibiru	4	Cipadung Cisurupan Palasari Pasirbiru
15	32.73.06	Cicendo	6	Arjuna Husen Sastranegara Pajajaran Pamoyanan ...
16	32.73.08	Cidadap	3	Ciumbuleuit Hegarmanah Ledeng
17	32.73.29	Cinambo	4	Babakan Penghulu Cisaranten Wetan Pakemitan Su...
18	32.73.02	Coblong	6	Cipaganti Dago Lebakgede Lebaksiliwangi Sadang...
19	32.73.27	Gedebage	4	Cimincrang Cisaranten Kidul Rancabolang Rancan...
20	32.73.16	Karacondong	7	Babakansari Babakansurabaya Cicaheum Compreng ...
21	32.73.13	Lengkong	7	Burangrang Cijagra Cikawao Lingkar Selatan Mal...
22	32.73.30	Mandalajati	4	Jatihandap Karangpamulang Pasir Impun Sindanglaya
23	32.73.28	Panyileukan	4	Cipadung Kidul Cipadung Kulon Cipadung Wetan M...
24	32.73.23	Rancasari	4	Cipamokolan Darwati Manjahlega Mekar Jaya
25	32.73.11	Regol	7	Ancol Balonggede Ciateul Cigereleng Ciseureuh ...
26	32.73.07	Sukajadi	5	Cipedes Pasteur Sukabungah Sukagalih Sukawarna
27	32.73.01	Sukasari	4	Gegerkalong Isola Sarijadi Sukarasa
28	32.73.19	Sumur Bandung	4	Babakanciamis Braga Kebonpisang Merdeka
29	32.73.26	Ujungberung	5	Cigending Pasanggrahan Pasirendah Pasirjati Pa...
30	NaN	TOTAL	151	NaN

2.2. Geo-coordinates of the districts in Bandung

Geo-coordinates of each district are readily available from the local government database: http://data.bandung.go.id/dataset/1fb076f0-24f8-4abe-b418-3b62083e15e8/resource/a6a429b4-18ff-4973-8310-c99ed4db6e08/download/koordinat-dan-ketinggian-kantor-kecamatan-di-kota-bandung-2014.csv

Panda's read_csv command was used to get the latitude and longitude data of each districts in Bandung.

```
[8]: lat_lon = pd.read_csv('http://data.bandung.go.id/dataset/1fb076f0-24f8-4abe-b418-3b62083e15e8/resourc
     lat_lon.head()
```

	Kecamatan	Lintang Selatan	Bujur Timur	Ketinggian (dpl)
0	Bandung Kulon	6.92520	107.5745	709
1	Babakan Ciparay	6.93710	107.5830	697
2	Bojongloa Kaler	6.93880	107.5938	694
3	Bojongloa Kidul	6.94285	107.5967	689
4	Astanaanyar	6.93190	107.5998	695

2.3. Top venues of districts

Top venues data were obtained from Foursquare through an API.

3. Methodology

3.1. Data cleaning

The list of districts obtained from Wikipedia were merged with geo-coordinates. Unnecessary features and NaN values were removed. This results in 28 districts found in Bandung with its latitude and longitudes exported to a dataframe, as shown in table below.

	Neighborhood	Latitude	Longitude
0	Andir	-6.90020	107.5740
1	Antapani	-6.90695	107.6646
2	Arcamanik	-6.91800	107.6850
3	Babakan Ciparay	-6.93710	107.5030
4	Bandung Kidul	-6.95900	107.6365
5	Bandung Kulon	-6.92520	107.5745
6	Bandung Wetan	-6.89728	107.6096
7	Batununggal	-6.93050	107.6371
8	Bojongloa Kaler	-6.93980	107.5938
9	Bojongloa Kidul	-6.94285	107.5967
10	Buahbatu	-6.96120	107.6607
11	Cibeurying Kaler	-6.90560	107.6313
12	Cibeurying Kidul	-6.90904	107.6371
13	Cibiru	-6.92910	107.7200
14	Cicendo	-6.90790	107.5974
15	Cidadap	-6.87330	107.5996
16	Cinambo	-6.92440	107.6952
17	Coblong	-6.88351	107.6110
18	Gedebage	-6.96260	107.6070
19	Kiaracondong	-6.09550	107.6770
20	Lengkong	-6.92910	107.6241
21	Mandalajati	-6.89750	107.6770
22	Panyileukan	-6.93600	107.6995
23	Rancasari	-6.94000	107.6749
24	Regol	-6.94607	107.6118
25	Sukajadi	-6.89269	107.5880
26	Sukasari	-6.86540	107.5916
27	Sumur Bandung	-6.91310	107.6179

Figure below shows the location of each districts, created using folium map by passing the coordinates of the district.

Foursquare API was used to obtain the venues of each district within 1km radius. There are total of 369 venues found in Bandung and they are summarised in the table below. The number of venues available at each district are found to be very contrast between one and another. More venues (more than 50) can be found in the districts located in the middle of Bandung city (i.e. 'Bandung Wetan', 'Cicendo', and 'Coblong'). There are two possible explanation. Firstly, number of venues at the district outskirt of the downtown are low in reality (i.e. not many venues available). Secondly, there is insufficient information recorded by Foursquare for the unpopular districts in Bandung (i.e. less people uploaded information/ data).

One-hot encoding was used to create dummy variables for each category to allow machine learning model to read and process the data. Top ten most common venues are then generated for each district and the data is ready to be fed into machine learning model.

3.2.   Machine Learning Model

K-means clustering model was considered to perform district clustering based on the 'similarity' of each district. Silhouette score was considered as the basis to determine the optimum number of clusters.

4. Discussion and Recommendations

Total of 3 clusters were considered based on the silhouette score above. The results of K-means clustering are shown in table below.

Hotel/Hostel	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue

Figure below depicts the cluster predicted by K-means.

Details of each cluster can be seen in the following table:

Cluster 1

Cluster 2

```
[47]: bandung_merged.loc[bandung_merged['Cluster Labels'] == 0, bandung_merged.columns[[0] + list(range(4, bandung_merged.shape[1]))]]
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
8	Bojongloa Kidul	Donut Shop	Pizza Place	Indonesian Restaurant	Yoga Studio	College Cafeteria	College Gym	College Stadium	Convenience Store	Cosmetics Shop	Cupcake Shop
13	Buahbatu	Donut Shop	Restaurant	Sports Club	Yoga Studio	College Auditorium	College Cafeteria	College Gym	College Stadium	Convenience Store	Cosmetics Shop

Cluster 3

```
[49]: bandung_merged.loc[bandung_merged['Cluster Labels'] == 2, bandung_merged.columns[[0] + list(range(4, bandung_merged.shape[1]))]]
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
13	Cibiru	Optical Shop	Soccer Stadium	Yoga Studio	Diner	College Auditorium	College Cafeteria	College Gym	College Stadium	Convenience Store	Cosmetics Shop

Cluster 1 is the most promising location to setup a business which comprises of various venues which could support the tourism industry. Proximity to the airport and hotels are one of the factors that Cluster 1 is the ideal location for tourism business.

5. Conclusion

Based on the results obtained from clustering, Cluster 1 is recommended as potential location to setup a business related to tourism (i.e. restaurant, hotel, etc)..

However, it is worth to note that the recommendation was drawn purely based on the availability of data in Foursquare database which might not be representative for certain area, as highlighted in Section 3 above.

It is also recommended to investigate other features such as population in each district, to narrow down/ pinpoint the best area for tourism business.